Likelihood (and Seals)

STAT 245

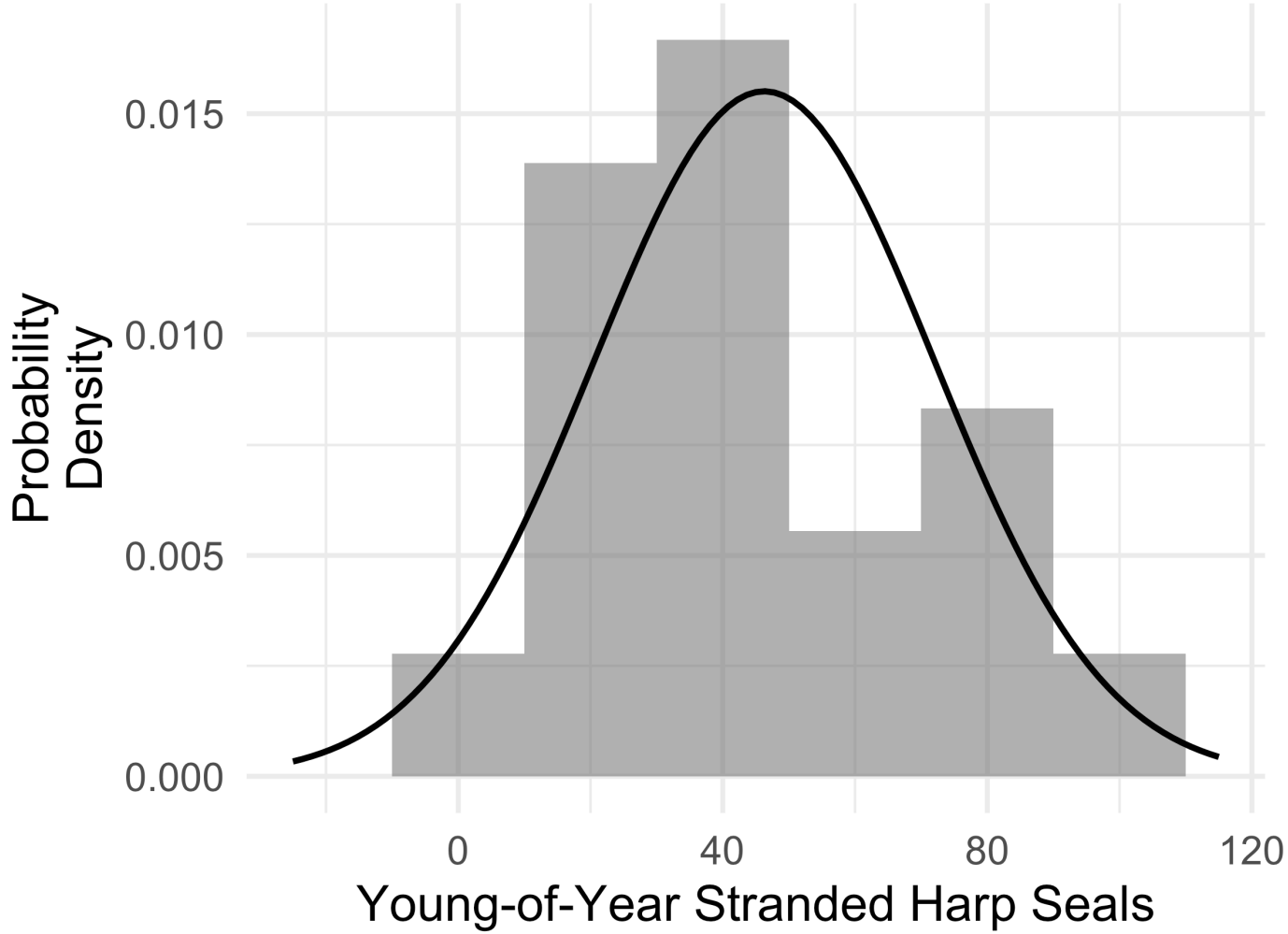
Harp Seal Strandings

Johnston *et al.*, "Effects of Climate Change on Harp Seals"

Harp seals use seasonal sea ice as a place to breed. Does climate change affect their survival? How many young-of-year are found stranded (dead) annually?

Tier Und Naturfotografie J und C Sohns / Getty Images

Annual Strandings



Normal probability density function (PDF)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A simple model

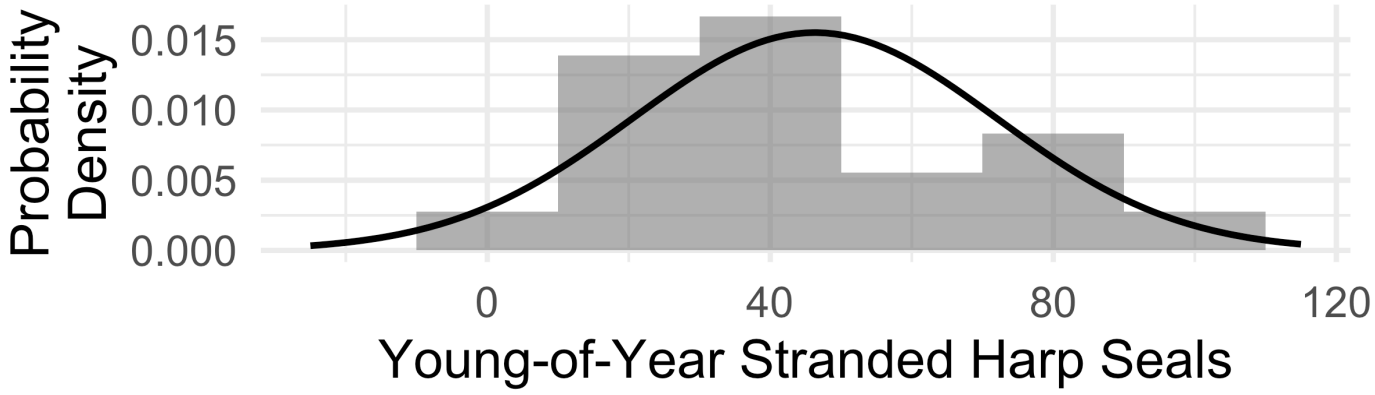
- Stranding distribution description:

```
df_stats(~strandings, data = harp_seals, mean, sd)
```

```
##     response     mean       sd
## 1 strandings 46.27334 26.47133
```

Model-based Predictions

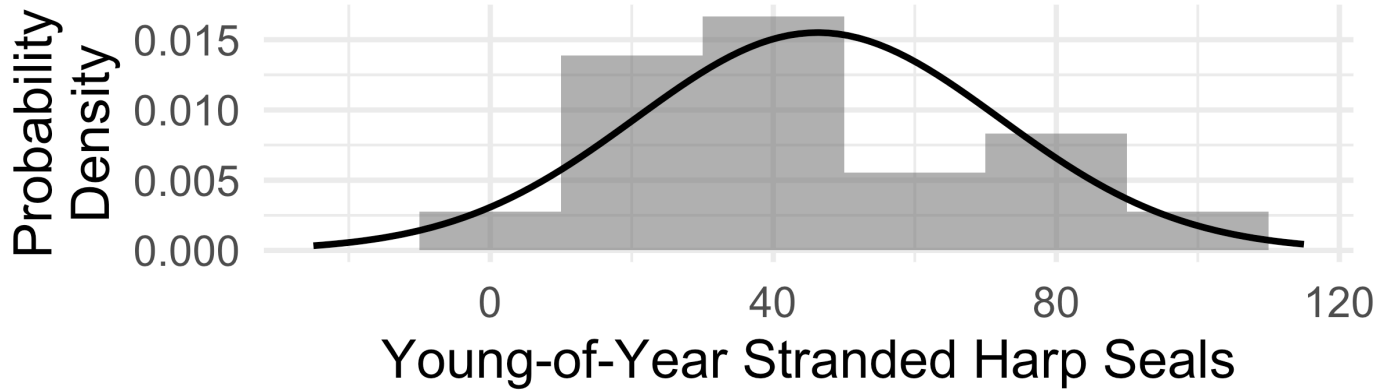
- Guess the number of strandings that will happen this year.
- Is it more likely that there will be 40 or 80 strandings?
- How much more likely?
- What is the *probability* of there being 47 strandings this year?



Likelihood to the Rescue!

Which is more likely?

- **Three years with 36, 41, and 43 strandings**
- Three years with 60, 41, and 40 strandings



Likelihood approach

- How did you:
  - Find the likelihood of each observation?
  - Combine the likelihoods of a set of three observations?
  - What did you have to assume about the set of observations?

How does all this seal stuff relate to linear regression?

Likelihood and `lm()`

How can we rewrite our model as a linear regression?

```
lm_version <- lm(strandings ~ 1,
                 data = harp_seals)
```

Model Equation

```
summary(lm_version)
```

```
##
## Call:
## lm(formula = strandings ~ 1, data = harp_seals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.481 -20.676  -1.391   7.150  47.737
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   46.273      6.239   7.416 1.01e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.47 on 17 degrees of freedom
```

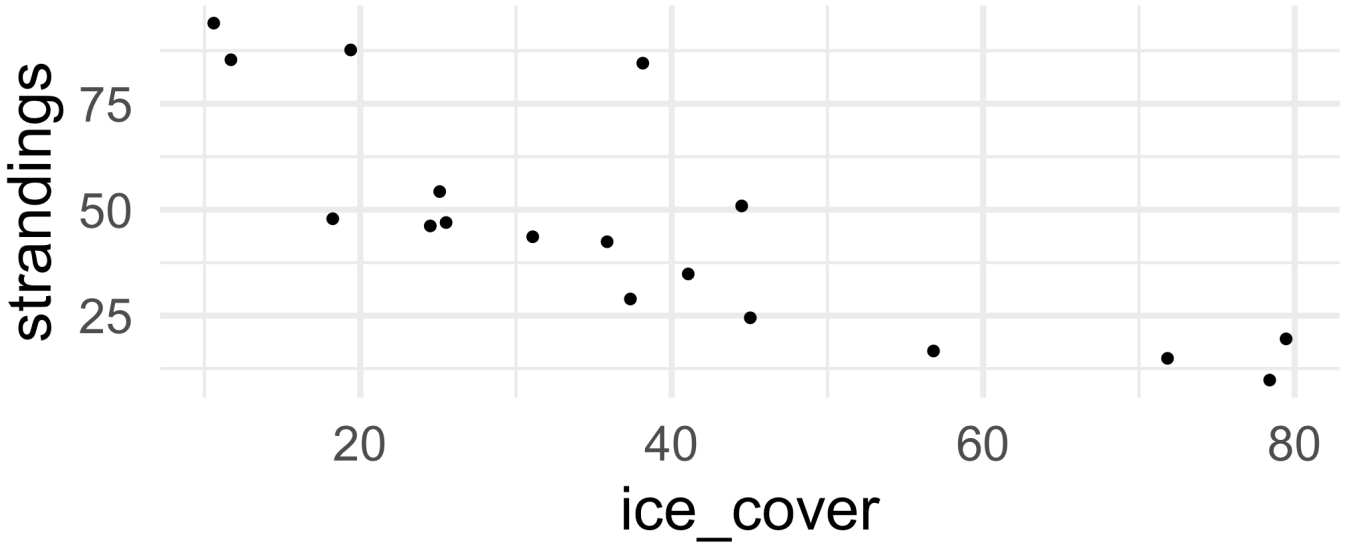
Compare models w/ Likelihood

it's likelihood *of the data, given a particular model*

Compare Models

w/Likelihood - Class example (teamwork problem)

```
gf_point(strandings ~ ice_cover,
         data = harp_seals)
```



```
lm_version_2 <- lm(strandings ~ ice_cover,
                                data = harp_seals)
msummary(lm_version_2)
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  84.2728     8.3397  10.105 2.37e-08
***
## ice_cover    -0.9848     0.1906  -5.168 9.33e-05
***
##
## Residual standard error: 16.7 on 16 degrees of
freedom
## Multiple R-squared:  0.6254,    Adjusted R-squared:
```

Model Comparison Challenge

joint likelihood of residuals given σ

```
harp_seals <- harp_seals |>
  mutate(resid1 = resid(lm_version),
         resid2 = resid(lm_version_2))
```

https://cutt.ly/seal-likelihood

https://www.danielsoper.com/statcalc/calculator.aspx?id=54

Likelihood...

- Can be used to measure model-data match
- (...and then as ingredient to AIC/BIC)
- What about *other* probability distributions?
- [Next...] Can be used to *fit* one model: which parameter estimates are "best"?